

本周工作：

● 语义项目：

讨论需要增加 NLP 模块。初衷是：当用户想要分析语义数据的时候，一开始并没有准确的目标，只想到几个关键词，比如餐厅，工作后去那些商场。我们需要根据用户输入的关键词，从我们的词向量空间中找到最有可能的轨迹（Topk 查询）。我们的方法是根据用户输入的关键词和 POI 词空间中的词进行比较，然后做轨迹查询。初步有一个大概的查询结果，用户来逐步地调整查询的输入。

可视化起到的作用有：

1. 帮助用户理解自己的输入，并优化
2. 帮助用户理解查询结果
3. 帮助用户理解为什么自己的输入导致了这样的查询结果，从而做出有益的输入调整。

根据符合输入的准确性得到轨迹的排序我想到的几个因素有：

1. 轨迹和输入的时空距离。（map 空间）
2. 轨迹和输入的语义空间（词向量空间）

所以我们允许用户输入的条件也从这两个空间下手。允许用户在 map 上画线或者词空间中调整线，定义自己的查询条件。这个 idea 进一步引申出了一个想法，轨迹分类器。轨迹分类器也可以做 VAUD 类似的那种节点，让用户逐步对大规模的轨迹进行分类，分类的方法有：

1. 选择算法进行分类
2. 选择空间条件划分轨迹
3. 选择时间进行划分
4. 选择语义空间中的空间条件进行划分

除了对某一个节点的轨迹进行分类，还能对节点的轨迹进行合并。从而形成交互式的轨迹分类器。直到找到自己想要的那部分轨迹，实际是一个过滤，合并的过程。还有一个创新点是轨迹的数据可以重载，而过滤的条件不变。

工程进展：

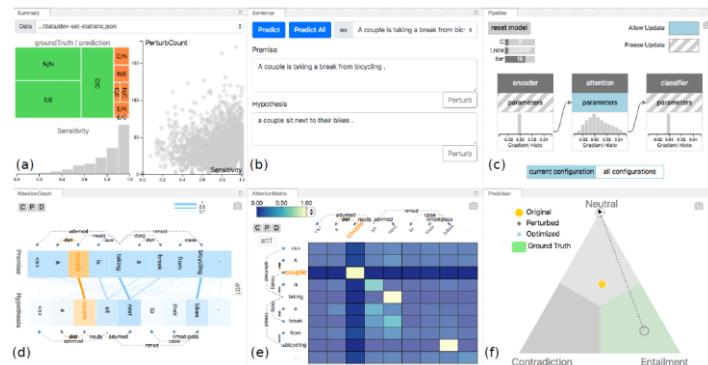
1. 实现了 POI 和输入的分词。输入的分词要分离出时空条件，这部分正在看 attention 机制，寻求解决办法。
2. 试了几个训练好的词空间库，他们的词都不能直接包含我们的 POI 分词，正在做加上 POI 分词的空间训练（直接调 API）
3. 轨迹分类器：实现了一个大规模轨迹的密度表达方法 视觉不会那么乱了 现在在做选择交互（基于 map 空间的条件选择和划分）

PS：这个轨迹分类器，应该不会做到这次的语义投稿中，和赵老师商量的结果是前端轨迹展示，只做简单可视化（这个密度表达是需要的，所以先做了）。

- 文章阅读：

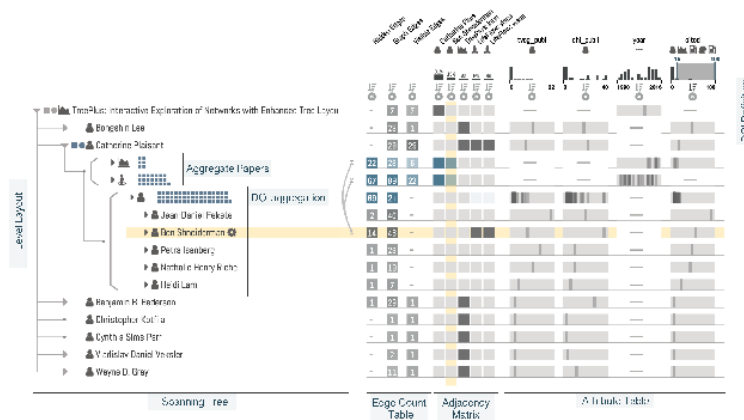
《NLIZE: A Perturbation-Driven Visual Interrogation Tool for Analyzing and Interpreting Natural Language Inference Models》

利用 attention 机制中的数值，帮助用户分析 NLP 模型中的过程，帮助用户理解 NLP 是怎么把输入分类变为输出，从而调整模型参数优化模型。其中下面矩阵和对应的图都是用于理解 Attention mechanisms。Attention mechanisms 是看输入的那一部分对输出影响比较大的一个量化方法。即重要的那部分



《Juniper: A Tree+Table Approach to Multivariate Graph Visualization》

把图结构，抽出来，抽成树结构，然后对每个节点就能展示其其他的属性值，并排序。简单，高效，实用。



《Embedded Merge & Split: Visual Adjustment of Data Grouping》

通过交互方法调整简单可视化表格的样式，形式等。本文的亮点在于聚焦在了 datagrouping 上而不是一种交互式调整的方法。写文章的同学值得借鉴。

《Drag and Track: A Direct Manipulation Interface for Contextualizing Data Instances within a Continuous Parameter Space》

本文提供了一种联动交互的方法，调整参数对数据进行（其实是）过滤的操作。

语义项目	NLP	完成分词，正在找匹配算法	
	语料库	加入 POI 分词训练中，之后应该会用二分图最大匹配计算距离	
	相似性度量		
	交互方法	密度表达方法和其加速。现在在做选择交互	
轨迹分类器	轨迹分类器	前段想法完善，没来的同学远程在实现轨迹的空间聚类	
	抽象空间轨迹	借鉴 visPoster 上雷达图的想法，看看能不能用多边形形成二维的多坐标轴空间，力引导布局，每个顶点是一个功能区。	
LMI	生成不同的图片		
流场项目（大屏）			